(19) **Japan Patent Office** (JP)

(11) Japanese Unexamined Patent Application Publication Number

(12) Japanese Unexamined Patent Application Publication (A)

S63-232626

(51) Int. Cl.[4]		Identification codes	JPO file numbers	(43) Publication date September 28, 1988
H 03 M	7/40		6832-5J	
G 06 F	3/06	301	W-6711-5B	
	5/00		H-7230-5B	
	12/04		C-8841-5B	

Request for examination Not yet requested Number of inventions 1 (Total of 3 pages)

(54) Title of the invention: DATA COMPRESSION AND DECOMPRESSION METHOD

(21) Japanese Patent Application: S62-66089

(22) Date of Application: March 20, 1987

(72) Inventor: Kimimasa SUZUKI, % Fujitsu Limited, 1015 Kamikodanaka, Nakahara-ku, Kawasaki-shi, Kanagawa-ken

(71) Applicant: Fujitsu Limited, 1015 Kamikodanaka, Nakahara-ku, Kawasaki-shi, Kanagawa-ken

(74) Agent: Patent attorney Takuichi IKEDA

SPECIFICATION

1. TITLE OF THE INVENTION

Data compression and decompression method

2. SCOPE OF PATENT CLAIMS

In a device that conducts Huffman compression and decompression of data using a compression table, a data compression and decompression method characterized in that it comprises a sampling means that obtains the current frequency of occurrence of characters and a reading and writing means that reads and writes character data in frequency of occurrence order.

3. DETAILED DESCRIPTION OF THE INVENTION

(ABSTRACT)

The present invention relates to the compression and decompression of data, and in order to improve data compression rates, it permits the table contents of the hitherto fixed compression table to be updated based on an analysis of the data that is the compression target, thereby providing the optimum compression for each compression target data.

(FIELD OF INDUSTRIAL APPLICATION)

The present invention relates to a data compression method, specifically Huffman compression.

Huffman compression is a technique for compressing character data into even shorter bit data, and it is generally used after a large volume of data has been character compressed. However, depending on the compression target data, the compression rate may deteriorate greatly, and a technique is needed to prevent this.

(PRIOR ART)

In Huffman compression, a compression table that assigns short bit strings in the order of characters that are thought to have high frequencies of occurrence is provided and data compression and decompression is conducted using this table.

(PROBLEM TO BE SOLVED BY THE INVENTION)

In conventional Huffman compression, characters are recorded in the order in which they are thought to generally occur most frequently. However, because this Huffman table is fixed, the method has the disadvantage that compression rates may be greatly degraded when the data contains large amounts of special characters.

(MEANS FOR SOLVING THE PROBLEM)

The Huffman compression method of the present invention is configured so as to comprise, in a device that conducts Huffman compression and decompression of data using a compression table, a sampling means that obtains the current frequency of occurrence of characters and a reading and writing means that reads and writes character data in frequency of occurrence order.

(OPERATION)

Figure 1 shows an example of an embodiment according to the present invention. In the figure, HSP is a Huffman compression and decompression section, HTBL is a Huffman compression table, FRM is a sampling means that obtains characters in frequency of occurrence order, and R/W is a reading and writing means that reads and writes data in frequency of occurrence order.

The Huffman compression and decompression section HSP comprises an input buffer IBUF that reads the data that is the target of compression or decompression, an output buffer OBUF that outputs the compressed or compressed data, and a conversion section HCDP that compresses or decompresses the data. The conversion section HCDP references the compression table HTBL and converts characters to bit strings or bit strings to characters. The compression table HTBL stores a character CHR and a corresponding bit string BIT that constitute a conversion pair. The processing above is the portion that is identical to conventional compression and decompression processing.

The sampling section FRM that is a feature of the present invention comprises a character selection section CSEL, a frequency of occurrence counting section CTR, a character buffer CBUF that stores occurring characters CHRA and the number of their occurrences, a sorting section SRT that sorts the characters in frequency of occurrence order, and a sorting buffer SBUF that stores the occurring characters in the order of their frequency of occurrence.

In the case of data compression, the sampling section FRM reads the data placed in the input buffer IBUF character by character and places each newly occurring character in the character field CHAR in the character buffer CBUF, and at this time, the counting section CTR sets a 1 in the frequency field FRQ that corresponds to that field in the character buffer CBUF as the frequency of occurrence. Thereafter, whenever a previously occurring character appears, the counting section CTR increases the frequency of occurrence. When the reading of all characters has been completed, the sorting section SRT sorts the characters in the order of the values in the frequency filed FRQ and outputs the characters to the sorting buffer SBUF. The reading and writing means R/W writes the group of characters in frequency of occurrence order to the compression table HTBL and also writes the compression data to the output buffer OBUF.

Thereafter, Huffman compression processing is executed by the conversion section HCDP.

In the case of data decompression, the reading and writing means R/W reads the character string data in frequency of occurrence order that has been written at the beginning of the data in the input buffer IBUF and writes it to the compression table HTBL. Thereafter, Huffman decompression processing is executed by the conversion section HCDP.

Figure 2 shows an example of an embodiment where the compressed data has been output to a magnetic tape. The group of characters that is stored in the Huffman compression table has been written in frequency of occurrence order in the data header HD and each blocked compressed data has been written to DA1, DA2, DA3, . . .

In this embodiment according to the present invention, the group of characters that is stored in the compression table has been written at the beginning of the compressed data, but it is not necessary that the group of characters that is stored in the compression table always be added to the beginning of the compressed data, and a means that sets the contents of the compression table from a separate compression occasion into the compression table at decompression may be used. This method is effective when the contents of the data change periodically.

In addition, because the units in which compression and decompression is conducted using the same compression table are unrelated to the volume of data, a completely different compression table may be used temporarily in the midst of a single compressed data. This method is effective when processing data formed of character group blocks that are extremely different.

(EFFECT OF THE INVENTION)

The present invention permits optimum Huffman compression to be conducted by setting a group of characters in frequency of occurrence order in a compression table for Huffman compression according to the data.

4. BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is an example of an embodiment of Huffman compression according to the present invention. Figure 2 is an example of an embodiment in which compressed data has been output to a magnetic tape.

In Figure 1, HSP is a Huffman compression and decompression section, HTBL is a Huffman com-

pression table, FRM is a sampling means, and R/W is a reading and writing means.

Patent attorney Takuichi IKEDA [seal illegible]

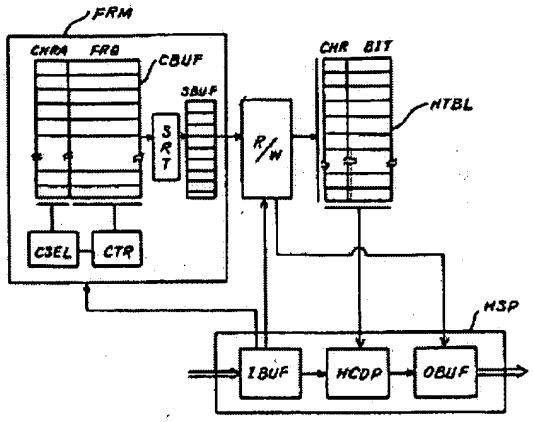

FIGURE 1

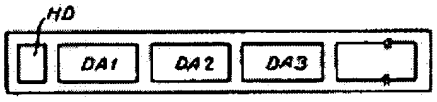

FIGURE 2